https://tecnohumanismo.online

TecnoHumanismo. Revista Científica
https://tecnohumanismo.online
Enero 2022
Volumen 2 / No. 1
ISSN: 2710-2394
pp. 114-125



Predicción de la contaminación atmosférica generada por las emisiones del CO_2 en el Perú utilizando los métodos ARIMA y Redes Neuronales

Prediction of atmospheric pollution generated by CO_2 emissions in Peru using ARIMA methods and Neural Networks

Previsão da poluição atmosférica gerada pelas emissões de CO_2 no Peru usando métodos ARIMA e Redes Neurais

ARTÍCULO GENERAL

Jaime Yelsin Rosales Malpartida
https://orcid.org/0000-0003-4574-5172
jrosalesm@uni.pe
Universidad Nacional de Ingeniería

Recibido 25 de Diciembre 2021 | Arbitrado y aceptado 25 de Diciembre 2021 | Publicado el 28 de Enero 2022

RESUMEN

El dióxido de carbono es el principal gas de efecto invernadero (GEI) que conduce al calentamiento global y, en consecuencia, al cambio climático y ambiental. Trae efectos negativos al desarrollo económico, la vida humana y el medio ambiente. Es de suma importancia poder medir y predecir de manera precisa la emisión de dióxido de carbono, ya que de esta manera podemos llevar a cabo una buena política sostenible para nuestro medio ambiente. Ante ello, la presente investigación tiene como principal objetivo encontrar el mejor modelo de predicción de las emisiones de dióxido de carbono (CO_2) en el Perú, mediante la evaluación comparativa de los métodos ARIMA y Redes Neuronales Artificiales. Se utilizó los datos anuales de las emisiones de CO_2 del Banco Mundial los cuales fueron analizados mediante la programación del software libre R studio. Para determinar el mejor modelo se utilizaron los errores de pronósticos tales como: raíz cuadrada del error cuadrático Medio (RMSE), error absoluto medio (MAE) y error porcentual absoluto medio (MAPE). Los resultados revelan que el modelo más apropiado entre estos dos métodos para la predicción de las emisiones del CO_2 en Perú es la red neuronal ANN(5-10-1), es decir, la red neuronal con cinco valores rezagados como entrada conectada por diez nodos en la capa oculta y una sola capa de salida, la cual tuvo una mayor precisión con un RMSE = 1125.82, MAE = 1040.68 y MAPE = 1.90 en la fase de prueba a comparación con el mejor modelo ARIMA(0,1,10) que tuvo un RMSE = 4223.73, MAE = 3143.40 y MAPE = 5.80 en la fase de prueba. En conclusión, se puede usar las redes neuronales para predecir las emisiones de CO_2, la cual mostró claramente que las emisiones anuales de CO_2 en el Perú aumentarán en los próximos años. Estos conocimientos reales serán útiles para que los responsables de la formulación de políticas traigan cambios significativos en las principales áreas medioambientales en nuestro país.

Descriptores: Modelo ARIMA, Redes Neuronales, GEI, CO_2.

ABSTRACT

Carbon dioxide is the main greenhouse gas (GHG) that leads to global warming and consequently to climate and environmental change. It brings negative effects to economic development, human life and the environment. It is of the utmost importance to be able to accurately measure and predict the emission of carbon dioxide, since in this way we can carry out a good sustainable policy for our environment. Given this, the main objective of this research is to find the best prediction model for carbon dioxide (CO_2) emissions in Peru, through the comparative evaluation of ARIMA methods and Artificial Neural Networks. The annual data of CO_2 emissions from the World Bank were used, which were analyzed by programming the free software R studio. To determine the best model, forecast errors such as: square root of the mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) were used. The results reveal that the most appropriate model between these two methods for the prediction of CO_2 emissions in Peru is the ANN(5-10-1) neural network, that is, the neural network with five lagged values as input connected by ten nodes in the hidden layer and a single output layer, which had a higher accuracy with RMSE = 1125.82, MAE = 1040.68 and MAPE = 1.90 in the test phase compared to the best model ARIMA(0,1,10) which had an RMSE = 4223.73, MAE = 3143.40 and MAPE = 5.80 in the test phase. In conclusion, neural networks can be used to predict CO_2 emissions, which clearly showed that the annual CO_2 emissions in Peru will increase in the coming years. These real insights will be useful to policymakers to bring about meaningful changes in major environmental areas in our country.

Descriptors: ARIMA Model, Neural Networks, GHG, CO_2.

RESUMO

O dióxido de carbono é o principal gás de efeito estufa (GEE) que leva ao aquecimento global e, consequentemente, às mudanças climáticas e ambientais. Traz efeitos negativos para o desenvolvimento econômico, a vida humana e o meio ambiente. É de extrema importância poder medir e prever com precisão a emissão de dióxido de carbono, pois assim podemos realizar uma boa política sustentável para o nosso meio ambiente. Diante disso, o objetivo principal desta pesquisa é encontrar o melhor modelo de previsão de emissões de dióxido de carbono (CO_2) no Peru, através da avaliação comparativa dos métodos ARIMA e Redes Neurais Artificiais. Foram utilizados dados anuais de emissões de CO_2 do Banco Mundial, que foram analisados por meio da programação do software livre R studio. Para determinar o melhor modelo, foram utilizados erros de previsão como: raiz quadrada do erro quadrático médio (RMSE), erro absoluto médio (MAE) e erro percentual médio absoluto (MAPE). Os resultados revelam que o modelo mais adequado entre esses dois métodos para a previsão de emissões de CO_2 no Peru é a rede neural ANN(5-10-1), ou seja, a rede neural com cinco valores defasados como entrada conectados por dez nós na camada oculta e uma única camada de saída, que teve maior precisão com RMSE = 1125,82, MAE = 1040,68 e MAPE = 1,90 na fase de teste em comparação com o melhor modelo ARIMA(0,1,10) que teve um RMSE = 4223,73, MAE = 3143,40 e MAPE = 5,80 na fase de teste. Em conclusão, as redes neurais podem ser usadas para prever as emissões de CO_2, o que mostrou claramente que as emissões anuais de CO_2 no Peru aumentarão nos próximos anos. Esses insights reais serão úteis aos formuladores de políticas para promover mudanças significativas nas principais áreas ambientais do nosso país.

Descritores: Modelo ARIMA, Redes Neurais, GEE, CO_2.

VISIT…

LANZAROTE
Caliente.com

1. Introducción

El cambio climático es uno de los principales problemas ambientales a nivel mundial, esto debido a las grandes emisiones de gases de efecto invernadero, principalmente dióxido de carbono (CO_2) [1]. Tras una caída sin precedentes del 5,4% en 2020, las emisiones mundiales de dióxido de carbono están volviendo a los niveles anteriores a la COVID y las concentraciones de GEI en la atmósfera siguen aumentando; para resolver el problema climático se requiere una reducción rápida y sostenida de las emisiones [2]. Las emisiones excesivas de CO_2 no solo son la causa del calentamiento global, sino también una amenaza para la supervivencia humana al causar problemas naturales y sociales como el cambio climático, el deshielo de los glaciares, el aumento del nivel del mar y la destrucción de la biodiversidad. Al realizar evaluaciones en Perú para determinar la vulnerabilidad y el riesgo en el sector agropecuario relacionado al clima, se encontró que, en la costa, los principales peligros climáticos son las inundaciones y sequías en las zonas bajas y en las partes altas la presencia de friajes y heladas, sumado a la presencia de eventos periódicos de El Niño. En la sierra, el 100% de las regiones sufren los impactos de las inundaciones y las heladas, mientras que el 80% reporta la presencia de friajes y sequías. En las zonas de selva o Amazonia peruana, el 100% de las regiones están expuestas a las inundaciones y friajes, mientras que el 85% de ellas también reportan la presencia de sequías [3]. Se demostró que existe una relación directa entre el crecimiento económico acoplado a las emisiones de CO_2 en Perú asimismo aumentan los GEI [4].

La Figura 1 muestra las emisiones de CO_2 en Perú por años (Banco Mundial, 2021). Como muestra la figura, las emisiones de CO_2 tienen una tendencia creciente. Las emisiones de CO_2, la principal fuente de emisiones de GEI, deben reducirse con el propósito de disminuir los efectos globales peligrosos.

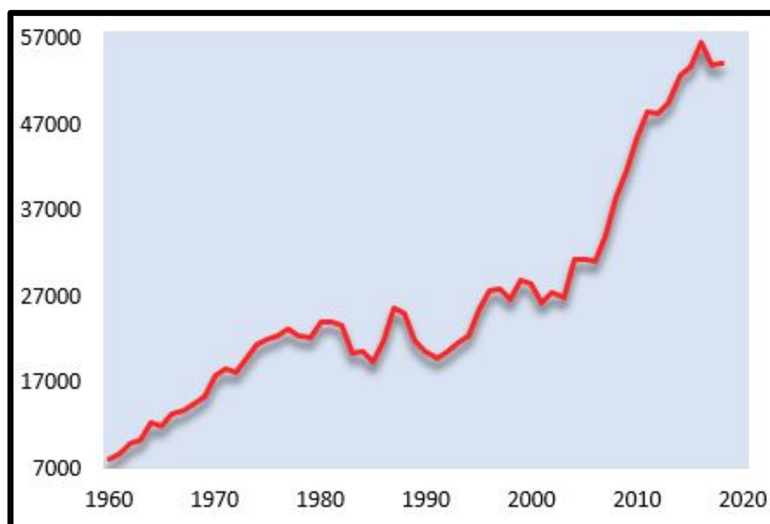


Figura 1. Emisiones anuales de CO_2 en Perú.

Se han realizado muchos estudios sobre los pronósticos de emisiones de CO_2 a través de series de tiempo. Longqi Ning, Lijun Pei y Feng Li [5] utilizó el modelo ARIMA para predecir las emisiones de CO_2 en China. Yawei Qi et al. [6] utilizó el modelo de redes neuronales para predecir la cuota regional de emisiones de CO_2. Wenhao Zhou et al. [7] utilizaron el método de modelado gris para predecir las emisiones de dióxido de carbono de China. Santos et al. [8] utilizaron una red bayesiana dinámica para el pronóstico emisiones de CO_2 en sistemas de generación de energía de múltiples fuentes en Brasil. Elham Shabani et al. [9] utilizó un enfoque novedoso para predecir las emisiones de CO_2 en el sector agrícola de Irán basado en un modelo múltiple inclusivo. Yi-Chung Hu et al. [10] utilizó un modelo optimizado de predicción fraccional gris para la predicción de emisiones de CO_2. Eyoh, Imo J, Umoeka, Ini J.1 y Udo, Edward N [11] utilizaron un sistema intuicionista de lógica difusa para la predicción global de emisiones de CO_2 en Nigeria. Egemen Hopali y Aslihan Cakmak [12] pronosticaron de manera diaria las emisiones de CO_2 de una fábrica de Turquía utilizando métodos estacionales ARIMA y Holt-Winters.

Por lo tanto, para este estudio, era importante comprender la trayectoria de emisiones de CO_2 en el pasado de Perú con el objetivo de hacer una predicción confiable de sus emisiones futuras con el fin de formular mejor las políticas y tomar medidas para reducir las emisiones de CO_2. El estudio pronosticó y modeló los niveles de emisión de CO_2 en Perú utilizando modelos ARIMA y Redes Neuronales.

2. Metodología

2.1. Modelo ARIMA y notaciones asociadas

El modelo econométrico ARIMA fue presentado por primera vez por Box & Jenkins en 1970 [13]. El modelo generalmente se ve favorecido por su flexibilidad para varios tipos de datos de series de tiempo y su precisión de predicción. ARIMA es una combinación de los modelos AR y MA, junto con la diferenciación. En los modelos autorregresivos (AR), las predicciones se basan en valores pasados de los datos de series de tiempo, y en los modelos de media móvil (MA), los residuos anteriores se consideran para pronosticar valores futuros. La forma general de ARIMA (p, d, q) se puede representar mediante un operador de desplazamiento hacia atrás como:

$$\Phi_p(B)(1-B)^d Z_t^{\lambda} = \Theta_q(B)\varepsilon_t \quad \ldots\ldots (1)$$

Donde los operadores característicos autorregresivos (AR) y promedio móvil (MA) son:

$$\Phi_p(B) = (1 - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p) \ \ldots\ldots (2)$$

$$\Theta_q(B) = (1 - \Theta_1 B - \Theta_2 B^2 - \cdots - \Theta_q B^q) \ \ldots\ldots (3)$$

Además, se cumple que:

$$(1-B)^d Z_t^{\lambda} = \nabla^d Z_t^{\lambda} \ \ldots\ldots (4)$$

Donde Φ es el parámetro estimado del componente autorregresivo, Θ es el parámetro estimado del componente de promedio móvil, λ es el parámetro de transformación por Box-Cox, ∇ es el operador de diferencia, d es la diferencia, B es el operador de rezago y ε_t es el término de perturbación.

2.2. La Metodología de Box - Jenkins

El primer paso hacia la selección del modelo es diferenciar la serie para lograr la estacionariedad.

Una vez finalizado este proceso, el investigador examinará el correlograma para decidir los órdenes apropiados de los componentes AR y MA. Es importante destacar el hecho de que este procedimiento (de elegir los componentes AR y MA) está sesgado hacia el uso del juicio personal porque no existen reglas claras sobre cómo decidir sobre los componentes AR y MA apropiados. Por lo tanto, la experiencia juega un papel

fundamental en este sentido. El siguiente paso es la estimación del modelo tentativo, después de lo cual se realizarán las pruebas de diagnóstico. La verificación de diagnóstico generalmente se realiza generando el conjunto de residuos y probando si satisfacen las características de un proceso de ruido blanco. De lo contrario, sería necesario volver a especificar el modelo y repetir el mismo proceso; esta vez desde la segunda etapa.

2.3. Recopilación de datos

El estudio se basó inicialmente en 59 observaciones de las emisiones totales anuales de CO_2 en el Perú (es decir, 1960-2018). Estos se obtuvieron de la base de datos del Banco Mundial actualizado al 2021, que es una fuente confiable de diversos datos macroeconómicos para todos los países del mundo. Tal es su credibilidad que muchos países realizaron este tipo de estudios utilizando esta base de datos para hacer sus investigaciones. Por lo tanto, la investigación tuvo que preferir esta fuente a base de su credibilidad y reconocimiento.

3. Resultados e interpretación

3.1. Análisis de los datos

Se procedió a dividir los datos de 1990 hasta 2014 para la fase de entrenamiento (Training o entrenamiento) y desde 2015 hasta 2018 para realizar nuestros pronósticos (Testing o prueba).

En primer lugar, para el modelo ARIMA se analizó los datos de predicción de emisión de CO_2 mediante pruebas de estacionariedad: análisis gráfico y estadística descriptiva.

3.2. Pruebas

Pruebas de diagnóstico y evaluación de modelos - Pruebas de estacionariedad: análisis de la Figura 1.

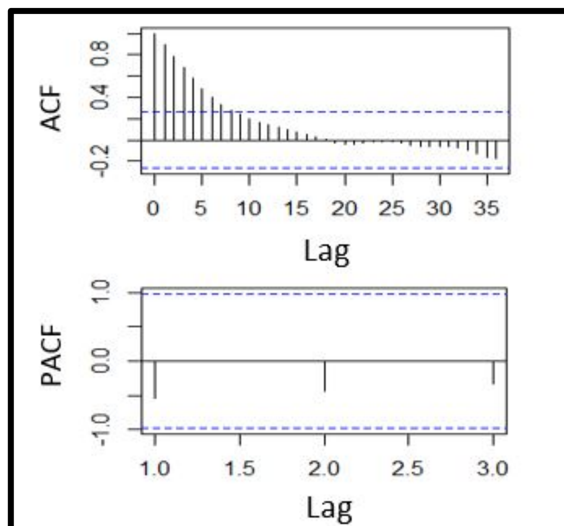


Figura 2. Gráficos de los correlogramas (ACF y PACF) para el modelo en fase de entrenamiento.

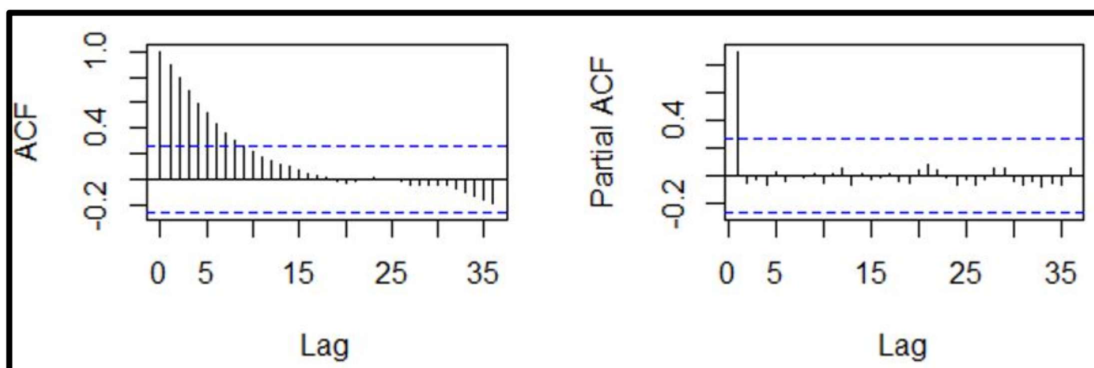


Figura 3. Gráficos de ACF y PACF para el modelo en fase de entrenamiento.

Como vemos que aún tiene una tendencia intensa (aun no es estacionaria) por ello procedemos a diferenciarla y vemos el gráfico realizando la primera diferenciación (d = 1).

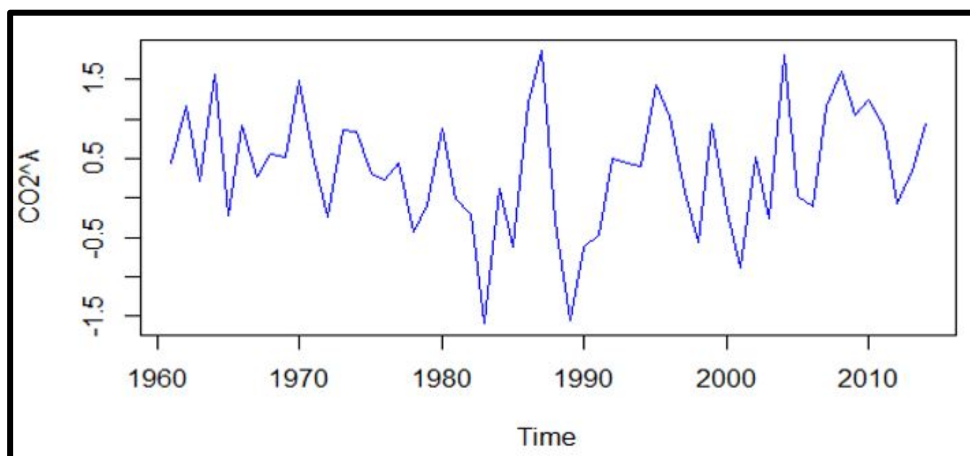


Figura 4. Gráfico de la serie 〖CO〗_2^λ diferenciada con d = 1.

Vemos nuevamente los correlogramas (Figura 5):

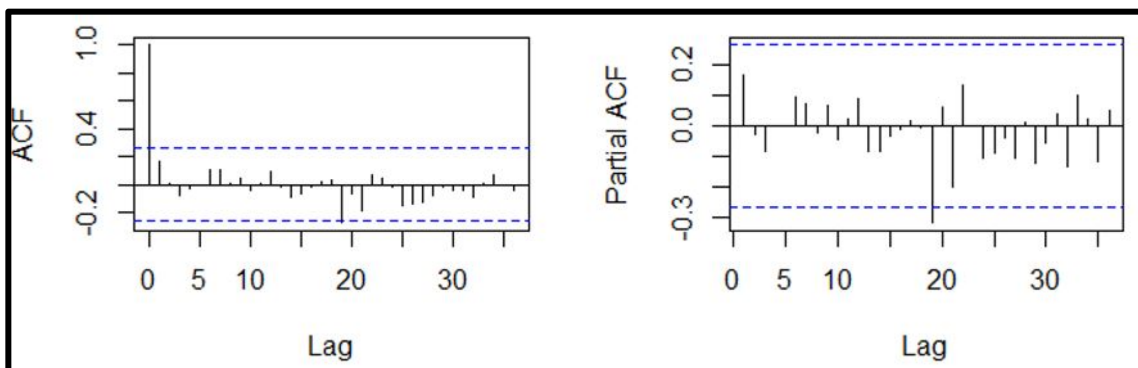


Figura 5. Gráficos de ACF y PACF para la serie 〖CO〗_2^λ en fase de entrenamiento.

Y por la prueba de Dick Fuller tenemos un p-value = 0.0402 < 0.05 la cual nos dice según la H0 que ya es estacionario y a partir de ello podemos identificar a nuestros modelos, las cuales en resumen son: ARIMA(1,1,8), ARIMA(1,1,7), ARIMA(1,1,6), ARIMA(7,1,1) y ARIMA(0,1,10); las cuales son los mejores modelos de un total de 30 modelos con distintos valores de p, d y q.

Tabla 1. Evaluación de modelos ARIMA.

Ajuste del modelo				
Modelo		RMSE	MAE	MAPE
ARIMA(1,1,8)	Training	1650.324	1304.579	5.41614
	Testing	6528.334	5406.255	9.950276
ARIMA(1,1,7)	Training	1651.282	1318.154	5.458692
	Testing	6341.742	5221.669	9.611984
ARIMA(1,1,6)	Training	1656.738	1315.751	5.450563
	Testing	6450.754	5356.532	9.857981
ARIMA(7,1,1)	Training	1669.879	1314.456	5.461572
	Testing	5913.827	4978.697	9.162236
ARIMA(0,1,10)	Training	1590.562	1110.441	4.758085
	Testing	4223.73	3143.398	5.796372

Tabla 2. Presentación de resultados del mejor modelo ARIMA(0,1,10).

Modelo	Parámetro	Coeficiente	Error Estándar	t- Calculado	P-Valor
ARIMA(0,1,10)	MA1	0.349	0.160	2.175	0.034
	MA2	0.101	0.168	0.602	0.549
	MA3	-0.139	0.158	-0.876	0.385
	MA4	0.192	0.156	1.236	0.221
	MA5	-0.020	0.160	-0.124	0.902
	MA6	0.115	0.131	0.876	0.385
	MA7	0.330	0.178	1.854	0.069
	MA8	-0.033	0.177	-0.186	0.853
	MA9	0.232	0.159	1.460	0.150
	MA10	0.182	0.181	1.008	0.318

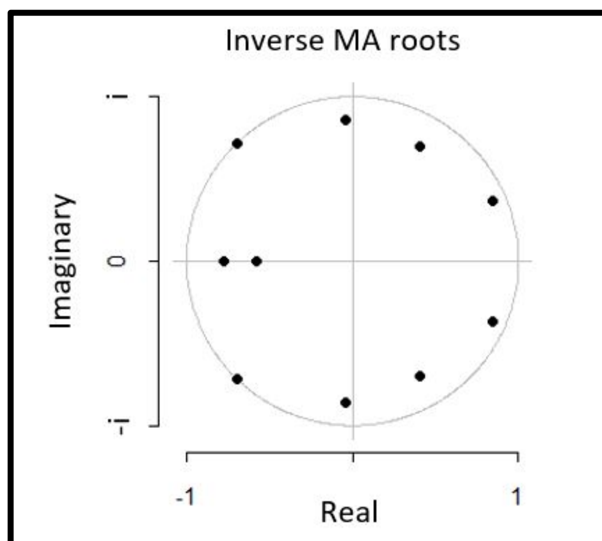


Figura 6. Prueba de estabilidad del ARIMA (0,1,10)

3.4.Hallazgos y discusiones

Tabla 3. Estadística descriptiva

Descripción	Estadístico
Media	26705.82
Mediana	23340.46
Mínimo	8173.74
Máximo	56530.00
Desviación Estándar	12565.63

Asimetría	0.96
Curtosis	3.13

De la tabla 3 tenemos que la media es positiva, es decir, 26705.81. La emisión mínima de dióxido de carbono es 8173.74 y la emisión máxima de dióxido de carbono es 56530. La asimetría es 0,96 y lo más importante es que es positiva, lo que demuestra que está sesgada positivamente y no es simétrica. La curtosis es 3.13 lo que indica que tenemos una distribución leptocúrtica y que la serie de emisiones de dióxido de carbono no se distribuye normalmente.

De la tabla 2 según los coeficientes estimados del mejor modelo ARIMA(0,1,10) la estructura de la ecuación para las emisiones de CO2 del Perú será la siguiente ecuación:

$$\nabla Z_t^{0.34} = 0.349\varepsilon_{t-1} + 0.101\varepsilon_{t-2} - 0.139\varepsilon_{t-3} + 0.192\varepsilon_{t-4} - 0.02\varepsilon_{t-5} + 0.115\varepsilon_{t-6} + 0.33\varepsilon_{t-7} - 0.033\varepsilon_{t-8} + 0.232\varepsilon_{t-9} + 0.182\varepsilon_{t-10}$$

3.5.Comparación de la Red Neuronal Feed-forward y ARIMA.

Al capturar el comportamiento complejo no lineal en una serie de datos, las redes neuronales son más efectivas y preferidas en lugar de los modelos ARIMA. Los resultados de la Tabla 4 revelaron que los datos de entrenamiento se ajustan utilizando la red de alimentación directa (Feedforward) 5-10-1, es decir, la red utiliza cinco valores rezagados (t-1, t-2, t-3, t-4, t-5) de series de tiempo como entrada, que está conectado por diez nodos ocultos a la capa oculta y conectado a una sola capa de salida. Para predecir las emisiones de CO2 (en la fase de entrenamiento), ANN predijo con valor RMSE = 707.13, MAE = 485.16 y MAPE = 1.96 y en la fase de prueba, se encontró que ANN predijo con valor RMSE = 1125.82, MAE = 1040.68 y MAPE = 1.90 mucho más preciso en comparación al mejor modelo ARIMA (0,1,10) que obtuvo un RMSE = 1590.56, MAE = 1110.44 y MAPE = 4.76 (en la fase de entrenamiento) y un RMSE = 4223.73, MAE = 3143.40 y MAPE = 5.80 (en la fase de prueba) no logrando superar a ANN durante las pruebas (Tabla 4) [14]. La precisión del pronóstico de los modelos de redes neuronales de alimentación hacia adelante fue consistente con respecto a RMSE, MAE y MAPE en las dos fases del ajuste del modelo, es decir, entrenamiento y prueba. Debido a que un modelo de red neuronal aprende de los datos, captura la no linealidad en la serie de datos de manera eficiente y finalmente, predice con mayor precisión que los modelos ARIMA. Por lo tanto, se prefirieron las redes neuronales en lugar de los modelos ARIMA para predecir las emisiones de CO2 de Perú.

Tabla 4. Presentación de resultados del mejor modelo ARIMA y Redes Neuronales.

Modelos	Criterio	Training	Testing
ARIMA(0,1,10)	RMSE	1590.56	4223.73
	MAE	1110.44	3143.40
	MAPE	4.76	5.80
ARIMA(7,1,1)	RMSE	1669.88	5913.83
	MAE	1314.46	4978.70
	MAPE	5.46	9.16
ANN(5-10-1)	RMSE	707.13	1125.82
	MAE	485.16	1040.68
	MAPE	1.96	1.90

La Figura 7 con un rango de pronóstico a partir de 2015 (líneas punteadas) hasta 2018 y la Tabla 4, muestran claramente que las emisiones totales anuales de CO2 de Perú aumenten durante los próximos años. El modelo con mayor precisión ANN (5,10,1) elegido aparentemente está enviando señales de advertencia a los economistas medioambientales de Perú sobre la necesidad de actuar a la luz del cambio climático y el calentamiento global.

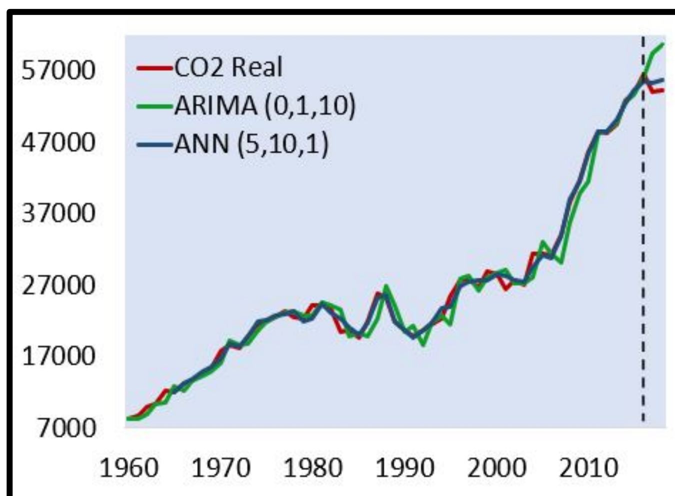


Figura 7. Rendimiento de predicción del mejor modelo ARIMA (0,1,10) y el mejor modelo de red neuronal Feed Forward ANN(5,10,1) para conjuntos de entrenamiento (1960-2014) y de prueba (2015-2018) de las emisiones de CO2 en Perú actualizado por el Banco Mundial al 2021.

4. Conclusión

El presente estudio comparó los modelos ARIMA con las redes neuronales artificiales, demostrando que el modelo de red neuronal ANN (5,10,1) es el modelo más adecuado para predecir el CO_2 total anual en el Perú durante los próximos años, dado que los modelos de redes neuronales están destinados principalmente a conjuntos de datos complejos no lineales y predicen de forma coherente cuando el conjunto de datos se divide en conjuntos de entrenamiento y de prueba. Según los resultados anteriores, la red neuronal converge a un ritmo más rápido a los mínimos locales y tiene la capacidad para analizar estructuras de datos complejas [6]. Mostrando claramente que las emisiones totales anuales de CO_2 de Perú aumenten en los próximos años. Estos conocimientos reales serán útiles para que los responsables de la formulación de políticas traigan cambios significativos en las principales áreas medioambientales en nuestro país.

5. Recomendaciones

- El gobierno de Perú debería reducir la contaminación implementando acciones políticas tales como aumentar los impuestos a las empresas contaminantes, especialmente a aquellas que utilizan combustibles fósiles en sus actividades de producción.
- Es necesario desarrollar o adquirir tecnologías de ahorro de energía más eficaces.
- Se recomienda el uso de energías renovables en Perú.
- Se recomienda esforzarse por educar continuamente a la sociedad de Perú sobre la esencia de la reducción de los niveles de contaminación.
- Se recomienda tener acceso gratuito a las mediciones del CO_2 de manera diaria, mensual o a tiempo real para futuras investigaciones y realizar modelos más eficientes que permitan tomar mejores decisiones.

Referencias

D. M. Maria, A. Adriana, L. Laura y J. Maritza, Rev. EIA, vol.18 (2021)

E. Hopali y A. Cakmak., Rev. Internacional de Información, Negocios y Gestión, vol. 12 (2020).

E. Shabani et al., Rev. Cleaner Production, vol. 279 (2021).

G. C. Mauricio David, tesis, Universidad Nacional de San Agustín de Arequipa, 2021.

J. Eyoh, I. Umoeka y, E. Udo, Rev. Internacional de Tendencias Avanzadas en Ciencias e Ingeniería de la Computación, vol. 10 (2021).

L. Ning, L. Pei and F. Li., Rev. Hindawi, Article ID 1441942 (2021).

L. O. Pedro, tesis, Universidad Nacional de Cajamarca PE, 2021.

S. R. Eduardo et al., Rev. JC, vol. 10 (2021).

Swaraj et al., Rev. Informática Biomédica, vol. 121 (2021).

T. M. Santos et al., IEEE International Systems Conference (SysCon), 2021, pp. 1-8.

UNEP, UNEP DTU Partnership, Informe sobre la Brecha de Emisiones, 2021.

W. Zhou et al., Rev, Chaos, Solitons & Fractals, vol. 147 (2021).

Y. Qi, W. Peng , R. Yan and G. Rao, Rev. Hindawi, Article ID 6659302 (2021).

Yi-Chung Hu et al., Rev. Internacional de investigación ambiental y salud pública, vol. 18 (2021).